MILESTONE

Project Acronym: Europeana Cloud

Grant Agreement number: 325091

Project Title: Europeana Cloud: Unlocking Europe's Research via The Cloud

MS21: Report with recommendations for enhancing EDM – first draft

MS21: Report with recommendations for enhancing EDM – first draft

Revision	Final version
Date of submission	
Author(s)	Valentine Charles, Europeana Foundation Marian Lefferts, CERL
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
V0.1	18-11-2014	Valentine Charles	Europeana Foundation	First draft of the document
V0.2	20-11-2014	Valentine Charles	Europeana Foundation	First draft revised by Marian Lefferts, CERL and Antoine Isaac, Europeana Foundation
V1.0	26-11-2014	Valentine Charles	Europeana Foundation	Final version

Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

The milestone MS21 is the first draft of the final deliverable *D4.4 Recommendations for enhancing EDM to support research oriented content* due at the end of the project (month 30).

The deliverable 4.4 aims at ensuring that the Europeana Data Model (EDM) is suitable for supporting the search and re-use of data and content by researchers within the infrastructure developed by the Europeana Cloud project. The deliverable should be based on the user needs identified in Task 1.2, the variety of content that will be placed in the Europeana Cloud (Task 4.2), requirements that may emerge from the metadata enhancements in Task 4.3 and the tools developed in WP3.

The task 4.4.1 was introduced in the Description of Work to ensure that the EDM was continually assessed against the needs of the project and its outputs. As stated in the Description of Work 'the aim is to ensure that EDM supports searching and re-use of data and content by researchers.' At the time of writing it was thought the emphasis would fall on refining or extending the existing EDM in relation to the metadata and content offered by the partners. However the first round of metadata aggregation from Europeana Cloud into Europeana has been done without triggering new requirements in terms of data modelling. The current EDM suited the Europeana Cloud metadata without requiring any amendments or additions to the model. Possible new requirements might emerge when data providers will start delivering content as planned for the end of the project. In addition the first outline of the Europeana Cloud architecture has been delivered month 10 and is currently in its early stages. It was therefore too premature to take these requirements into account for the writing of this milestone.

It is for these reasons that WP4 would therefore like to change the scope of the deliverable 4.4 and would like to use this milestone to examine two issues that are highly relevant for the future development of the Cloud-based aggregation infrastructure the project is putting in place.

These two directions are proposed considering the current state of the Europeana Cloud architecture:

- Extend the EDM to support the ingestion of content in Europeana Cloud

- Extend the EDM to model the data flows defined within the Europeana Cloud infrastructure

Extend the EDM to support the ingestion of content in Europeana Cloud

One of the objectives of Europeana Cloud is to demonstrate the ingestion of content within the Cloud infrastructure. The completed technical infrastructure as developed in the final deliverable *D2.6) Metadata and Content Cloud Delivered: Metadata and Content Cloud Delivered - Functional Release v1.0; associated report on the Evaluations of the Metadata and Content Cloud Performance* will highlight new requirements that might affect the provision of content to Europeana using EDM.

4

The deliverable 4.4 will therefore address more specific data modelling requirements related to the delivery of content in Europeana as planned by the deliverable *4.5)Research metadata and content available in the Europeana Cloud*. At the time of writing the work on ingesting content and on its supporting workflow has just started (with the bulk of the work scheduled in M24-36), and potentially new modelling requirements for EDM are beginning to emerge. The ingestion of content such as images, videos, text files and full-text will indeed require from the EDM to support a new set of technical metadata describing the digital representation of a given cultural heritage object. The extension of EDM with additional technical metadata will contribute to the creation of the new Europeana Content Re-Use Framework[1]. The work on the definition of the Content Re-Use framework has started within the Europeana Creative[2] project and will be pursued within Europeana Cloud as well. It creates a new set of conditions to facilitate the access and re-use of the Europeana metadata and content.

Europeana Cloud will ingest content related to two categories:

- Digital representation of a given cultural heritage objects such as scanned pages of a book, a thesis.
- Born-digital objects such as videos, OCRed texts, or e-books.

Depending on the technical requirements of the new architecture, D4.4 will work on the implications of these requirements on EDM. Work done by other projects such as Europeana Libraries[3] will serve as a basis for D4.4.

Extend the EDM to model the data flows defined within the Europeana Cloud infrastructure

The first outline of the Cloud architecture as described in deliverable *2.2) Initial Version of Architectural Design Document: Initial version of the Architectural Design Document* has formulated initial requirements in terms of data flows within the Europeana Cloud. The figure below taken from the deliverable D2.2[4] shows the different transactions which could take place within the Cloud:

- A data provider A will upload its original data to the Europeana Cloud.
- It will map it to a different format, for instance EDM for provision to Europeana.
- It will regularly update its uploaded datasets.
- A data provider B will export the data from data provider A.
- It will enriched the data from data provider A and will re-upload it the Cloud and so on.
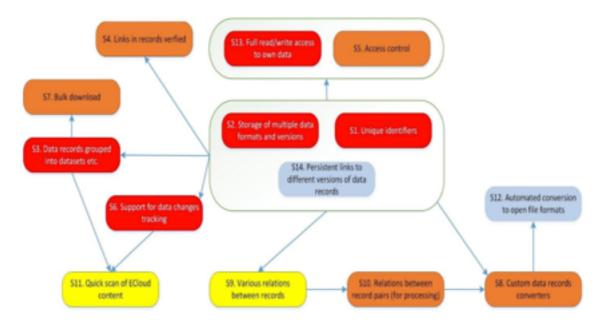
---

[1] http://pro.europeana.eu/web/europeana-creative/extended-europeana-licensing-framework

[2] http://pro.europeana.eu/web/europeana-creative/home

[3] D4.3 Report on how the full-text content will be made available to Europeana: http://www.europeana-libraries.eu/documents/868553/c52bacb5-f810-49d3-8e8a-a87e00da9a29

[4] Figure 2. Initial dependencies between summarised user stories for Europeana Cloud system with three stages of development: ALPHA (red), BETA (orange) and 1.0 (yellow).

Most of these transactions will need to be captured in terms of metadata for managing the resources.

We can already highlight few steps within the Cloud architecture that could be expressed using extensions of EDM:

Capture metadata at dataset level

The Europeana Cloud system will unable the storing, distribution and re-use of cultural data. Digital files representing cultural heritage objects and their metadata descriptions will be managed as datasets. Each dataset will have provenance information, versions, different formats that will need to be captured as metadata. Any changes applied to the metadata should be monitored at the level of the dataset. Europeana has already initiated some work in this area and has published a EDM dataset profile[5] that could be used as a basis for further work in D4.4.

Capture information about data providers

In addition to metadata at the dataset level, each data provider should be described within the Europeana Cloud architecture with information such as the name of its institution, its country, contact details etc. These metadata will be used for administrative purposes but also to track the provenance of the data within the different data workflows the Cloud enables.

---

[5] EDM dataset profile http://pro.europeana.eu/documents/900548/3345fdd3-9d81-4828-a1f6-da099a394e4e

The EDM profile for organisation[6] could be use as a basis for D4.4. The flexibility of EDM enable extensions of this profile if needed.

Representation of technical metadata for the uploaded content

As described earlier in this document, the Europeana Cloud infrastructure will enable the storage and re-use of digital files representing cultural heritage objects. Re-users of the content will need to know the technical specificities of these files. These characteristics will need to be capture in a new set of EDM properties. The Europeana Sound project has for instance define a set of technical metadata in its EDM profile for Sound[7] that could be re-used in the context of Europeana Cloud.

Representation of data enrichment

The Europeana Cloud infrastructure will allow the re-use of metadata by many stakeholders including Europeana. Metadata will be possibly enriched by different parties before being re-uploaded in the Cloud. It is therefore important to document the enrichment performed at different stages of the process and by different parties. EDM could be extended for this purpose.

WP4 will closely follow the work within WP2 and will decide where to focus its efforts for the final deliverable 4.4. WP4 will also take its decision based on the relevance of the work for the whole Europeana Network. The deliverable will probably get a different title than initially planned in the description of work that will reflect the decisions made with WP4.

---

[6] EDM organisation profile http://pro.europeana.eu/documents/900548/2f1b6cfd-b803-4f01-8128-6749b59cd2da

[7] EDM profile for Sound http://pro.europeana.eu/documents/2011409/5c845fc2-dcf7-46c2-a66d-6aea9dac1c0c